The icing on the cake - combining relational and semantic methods to extract meaning from online message board postings

Moser, C.; Hellsten, I.; Groenewegen, P.

published in
Proceedings of the ACM WebSci'11
2011

document version
Publisher's PDF, also known as Version of record

The Icing on the Cake – Combining Relational and Semantic Methods to Extract Meaning from Online Message Board Postings

Christine Moser
VU University Amsterdam
De Boelelaan 1081
NL-1081 HV Amsterdam
+31 20 5989149
c.moser@vu.nl

Iina Hellsten
VU University Amsterdam
De Boelelaan 1081
NL-1081 HV Amsterdam
+31 20 5986983
i.r.hellsten@vu.nl

Peter Groenewegen
VU University Amsterdam
1st line of address
2nd line of address
+31 20 5986914
p.groenewegen@vu.nl

ABSTRACT
In this paper, we propose to combine relational and semantic approaches toward studying online communities. In particular, we demonstrate how to integrate social network analysis data and semantic maps, both extracted from one data set. Integration yields insights that lead to more than just a combination of insights. Specifically, we find that individuals occupying central network positions use distinct word configurations and are concerned with very different topics. Thus, we provide a relational network with meaning, and discuss how this finding might be employed in future research.

General Terms
Measurement, Documentation, Human Factors, Languages, Theory.

Keywords
Online Communities; Social Network Analysis; Semantic Maps; Content Analysis; Mixed Methods.

1. INTRODUCTION
In recent years, we have seen a rapid growth of online communities on the Web. These virtual communities serve as socio-technical platforms for various professionals, entrepreneurs and serious hobbyists to engage in discussion around a shared area of interest. They mostly use these platforms to exchange knowledge and expertise. So far, online communities have been analysed either via the structure of the communities or the content of messages and the motivations of members. We argue that in order to gain insight in the dynamics of online communities, we have to combine a set of methods that allows for the analysis of both the structure as well as the content of communications in these communities. Empirical studies in the domain of online communities usually employ a single method. Often, motivations to participate in these communities were investigated [e.g., 1, 2]. Other research focused on behaviour of community members [e.g., 3, 4]. Some of these studies employed qualitative methods, especially case studies [5-6] and ethnographies [7-8]. However, studies that combine different methods are scarce [9]. This holds in particular for studies that focus on relational (using social network analysis) and interpretational information (using for example semantic maps). To our knowledge, studies that employ both methods are lacking. With this paper, we want to contribute to the literature by proposing a way to combine both approaches. We illustrate the approach with data from an online community, and discuss implications for researchers.

WebSci '11, June 14-17, 2011, Koblenz, Germany.

2. EARLIER RESEARCH AND METHODOLOGICAL POSITIONING
Recently, several approaches to combining network structure and network content have been conducted in information science, computer sciences and bibliometrics. Danowski and Cepela [10] have conducted automated network analysis of social actors using text corpora. Mapping the co-actor networks is similar to the basic bibliometrics analysis of co-authorships in scientific publication that has flourished in bibliometrics and scientometrics since the 1980s. Ereteo et al. [11] have specifically focused on combining the structure and content of networks by manually adding tags to messages. Interestingly, their results suggest that the semantics in the networks affect the network structure. There are also efforts to develop content based social network analysis [12]. Their main interest is in analyzing shared topics between groups. In a similar vein, Roth and Cointet [13] have studied the co-evolution of social and socio-semantic networks on large-scale networks of scientists (co-authorships and co-topics), and between bloggers (hyperlinks and topics). Xu and colleagues [14] have combined social networks with semantic concepts analysis by comparing the networks of researchers to the networks of concept similarity. Another step toward the integration of interpretive and statistical methods has been undertaken recently, showing that social network analysis and discourse analysis can be usefully integrated [15]. However, in our analysis we will pay attention to how semantic co-word maps are related to the network structure. This approach combines two automated analytical processes, focused on individual actors in a network. Thus, this approach differs from earlier approaches that combine qualitative and statistical methods by integrating automated analytical processes. Furthermore, this method allows us to zoom in on individual actors, both in terms of their structural positioning in the network, as well as the content of their individual messages, and thus offers new perspectives that build on abovementioned studies.

3. SETTINGS AND METHODS

In order to make our point, we conduct an illustrative study of members of one online community of Dutch cake bakers with about 13,000 members. First, we conducted a social network analysis (SNA) to analyze the networked structure of the communications in the community. We selected data from one particular day that showed a lot of activity and promised to yield good results. In total, 212 actors participated in discussion on that day, leaving 1378 messages. We then compiled an affiliation matrix: actors were entered in rows, events (defined as topics where messages were left) in columns. We defined a network tie as simultaneous presence of messages by individual community members at the same topic. This affiliation matrix was then transformed into a bipartite graph [16] which allowed us to calculate member centrality measures using Ucinet 6 [17].

For the next analytical step, we investigated all messages of the three most central actors, who posted 67, 85 and 129 messages, respectively. We used a co-word based semantic maps approach that shows the implicit frames within the discussions [18]. The method automatically maps the positions of words in a set of documents, on the basis of asymmetrical word-document matrices where documents form the rows and the words the columns. Salton's index (cosine) is used for the normalization. The positions of words in a set of documents can be considered as the unintended results of a set of relations in a network among agents or documents [19-20]. The method focuses not only on dyadic co-occurrences of words but also triadic etc. co-occurrences. For the visualization we did not use the co-occurrence matrix but the underlying asymmetrical matrix of documents versus words, and subsequently computed the distance among the word vectors using the vector-space model, that is, using the cosine as a similarity measure [21].

4. FINDINGS

The first part of our findings consists of centrality measures of all actors within the social network. In particular, we focused on degree centrality. Table 1 shows that the three most central actors indeed occupy the top with quite some distance to the rest of the actors[1]. As these three actors prove to be most active in this network, further investigation of their results should yield interesting results.

Table 1. Degree centrality of top 20 actors

Actor	Normalized degree centrality
1	7.916
2	6.406
3	5.287
4	4.502
5	4.154
6	3.836
7	3.599
8	3.406
9	3.318
10	3.118
11	2.784
12	2.777
13	2.747
14	2.555
15	2.540
16	2.481
17	2.325
18	2.155
19	2.066
20	1.985

[1] We only show the top 20 actors, to give an impression of the differences between actors.

Next, we investigated the semantics of the top three actors' messages using automated mapping of the positions of words in the documents of the three most central posters, Actor 1, Actor 2 and Actor 3, respectively, as mentioned in Table 1. Due to the relatively larger number of messages of actor 3, we present a semantic map of this actor that contains unique words used three times or more, in order to be able to present a clearly arranged map. The other two semantic maps present unique words that occur twice or more. Colours of nodes in the semantic maps indicate different word clusters, whereas their size relates to the frequency of words. Finally, tie strength indicates the frequency of relations between words.

The semantic map of Actor 1's messages mainly contains admiring comments, such as 'looks just super', 'what a beautiful cake', and supporting comments like 'well done', and 'want to bake that this week as well...'. In the semantic map, words such as *good*, *great*, *delicious*, *wow* and *ooow* occur often as well as the smiley *big smile*, and *smiley lol* (Figure 1). The semantic map is therefore quite coherent.

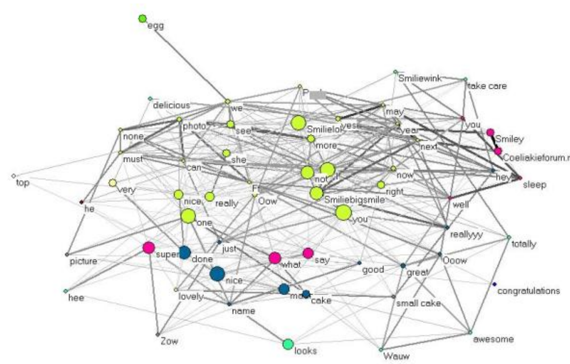

Figure 1. Semantic map of actor 1's messages; 61 unique words (occurring twice or more often, stopwords removed), 67 postings, cosine: 0.303

The messages of Actor 2 are more diverse in their semantics. They show support, but on a slightly more moderate level, using expressions such as 'what a nice one girl', 'can imagine you want to try everything now', 'unfortunately the photo is not quite as I wanted…’. There are more smileys in use than in Actor 1’s messages, covering the whole range from *smiley big smile, smiley yummie* to *smiley huh* (Figure 2). As such, this semantic map shows more variation than that of Actor 1, and emphasizes the use of smileys.

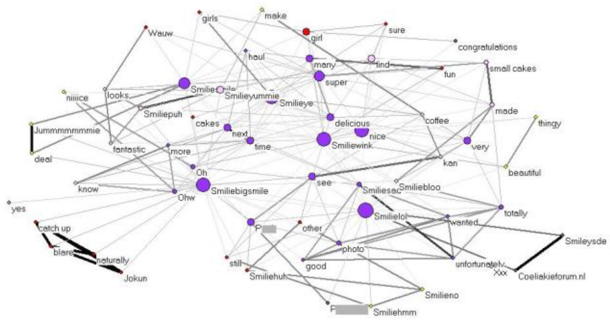

Figure 2. Semantic map of actor 2’s messages; 62 unique words (occurring twice or more often, stopwords removed), 85 postings, cosine 0.302.

Actor 3 posted more messages than the more central actors 1 and 2 during our period of analysis. Most messages concern heading for a journey (hotel, room, airplane, airport, and suitcase). There are also several postings that are probably meant for insiders, such as ‘well, I can’t do anything about it that you are jealous’, or ‘there are things I just cannot post here in the forum *smileylol’* (Figure 3). Overall, her semantic map is much more diverse in composition than the ones from Actor 1 and 2.

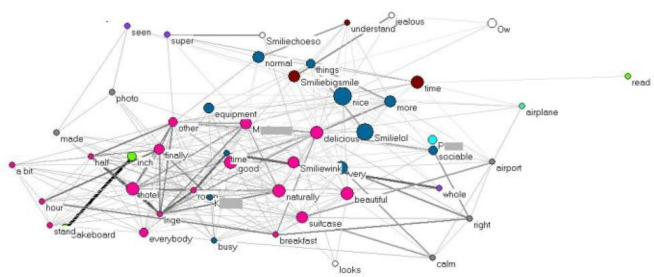

Figure 3. Semantic map of actor 3’s messages; 49 unique words (words occurring three times or more, stopwords removed), 129 postings, cosine 0.265

When comparing the three semantic maps, several differences stand out. Actor 1’s words produce a quite coherent network, which mainly consists of compliments and support for others. The actual content of her messages remain unclear, she seems to focus mostly on other actors’ messages. Actor 2’s messages lead to a slightly different picture: although this semantic map is also quite coherent, there is more variation (hence a larger semantic map). Whereas she also seems to compliment others, she extensively uses smileys in her messages. Finally, Actor 3 uses a wider scale of words, in particular words related to cakes. This leads to a more varied semantic map. Apparently, her use of words is less coherent than Actor 1’s and 2’s.

5. DISCUSSION AND CONCLUSION

Our results show that individuals who occupy central positions in a social network use distinct configurations of words .Whereas the first investigated member dominantly uses words that serve as compliments and support for others, the second member extensively employs smiley’s to make her point. Finally, the third member uses a wider scale of words, in particular words related to cakes. Thus, although the three actors all occupy central positions in the social network, their use of words is significantly different. This is interesting, because it shows that individuals feature different styles when composing messages, and that their messages are concerned with very different content. From an institutional theory perspective [e.g., 22], the question arises how, from a micro-lens, this is possible, considering that this particular community is very cohesive and features strong social norms. Future research might investigate the social mechanisms that both enable and restrain individuals in their use of words when composing messages.

Our study shows that a relational and semantic approach to analyzing online messages yields rich results. Not only is it possible to shed light on the relational structure of a network, such as the position of different actors. We also are able to investigate the content of messages, using exactly the same data. This content analysis provides the network with meaning: it informs us about what flows between different actors, in addition to knowledge about their connectedness.

This study contributes to the emerging literature arguing for more use of cross-disciplinary methodology [e.g., 23]. In particular, it provides a point of departure for studies into social networks that aim to not only map out their social structure, but also want to attach meaning to this structure. Thus, our approach is a first step toward a more encompassing method in social network research.

Finally, we open up new perspectives for scholars who are mainly concerned with the analysis of text. The rise of Internet technology, with its accompanying rush of social media and other communication outlets, provides scholars with new possibilities concerning data collection and analysis. As we show in this article, the same dataset provides us with the possibility to perform different kinds of analyses, such as content analysis and social network analysis. Scholars who predominantly perform content analysis might want to consider extending their approach with other possibilities, such as social network analysis.

6. REFERENCES

[1] Hertel, G., S. Niedner, and S. Herrmann, *Motivation of software developers in Open Source projects: an Internet-based survey of contributors to the Linux kernel.* Research Policy, 2003. **32**: p. 1159-1177.

[2] Jeppesen, L.B. and L. Frederiksen, *Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments.* Organization Science, 2006. **17**(1): p. 45-63.

[3] Lakhani, K.R. and E. von Hippel, *How open source software works:“free” user-to-user assistance.* Research Policy, 2003. **32**(6): p. 923-943.

[4] Franke, N. and S. Shah, *How Communities Support Innovative Activities: An Exploration of Assistance and Sharing Among End-Users.* Research Policy, 2003. **32**(1): p. 157-178.

[5] Piller, F., et al., *Overcoming Mass Confusion: Collaborative Customer Co-Design in Online Communities.* Journal of Computer-Mediated Communication, 2005. **10**(4): p. article 8.

[6] Ross, D.A.R., *Backstage with the Knowledge Boys and Girls: Goffman and Distributed Agency in an Organic Online Community.* Organization Studies, 2007. **28**(3): p. 307-325.

[7] Füller, J., G. Jawecki, and H. Muhlbacher, *Innovation creation by online basketball communities.* Journal of Business Research, 2007. **60**(1): p. 60-71.

[8] O'Mahony, S., *Guarding the commons: how community managed software projects protect their work.* Research Policy, 2003. **32**: p. 1179-1198.

[9] Wasko, M. and S. Faraj, *Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice.* MIS Quarterly, 2005. **29**(1): p. 33-57.

[10] Danowski, J. and N. Cepela, *Automatic mapping of social networks of actors from text corpora: Time series analysis.* Annals of Information Systems: Data Mining for Social Network Data (Memon, N. et al. eds), 2010. **12**.

[11] Ereto, G., et al., *Semantic Social Network Analysis, a concrete case*, in *Handbook of Research on Mehtods and Techniques for Studying Virtual Communities: Paradigms and Phenomena*. 2011, IGI Global: Hershey, PA.

[12] Velardi, P., et al. *A new content-based model for social network analysis*. in *IEEE International Conference on Semantic Computing*. 2008.

[13] Roth, C. and J.-P. Cointet, *Social and semantic coevolution in knowledge networks.* Social Networks, 2009. **32**: p. 16-29.

[14] Xu, B., R.R. Jones, and B. Shao, *Volunteers' involvement in online community based software development.* Information & Management, 2009. **46**: p. 151-158.

[15] Moser, C., P. Groenewegen, and M. Huysman, *Extending Social Network Analysis with Discourse Analysis – Combining Relational with Interpretive Data*, in *The Influence of Technology on Social Network Analysis and Mining*, T.e.a. Ozyer, Editor. forthcoming, Springer: New York.

[16] Faust, K., *Centrality in affiliation networks* Social Networks, 1997. **19**: p. 157-191.

[17] Borgatti, S., M.G. Everett, and L.C. Freeman, *Ucinet for Windows: Software for Social Network Analysis*. 2002, Harvard, MA: Analytic Technologies.

[18] Hellsten, I., L. Leydesdorff, and J. Dawson, *Implicit media frames: Automated analysis of public debate on artificial sweeteners.* Public Understanding of Science, forthcoming.

[19] Leydesdorff, L. and I. Hellsten, *Measuring the meanings of words in contexts: automated analysis of 'Monarch butterflies', 'Frankenfoods' and 'stem cells'.* Scientometrics, 2006. **67**(2): p. 231-258.

[20] Burt, R.S., *Toward a structural theory of action: network models of social structure, perception, and action*. 1982: Academic Press.

[21] Leydesdorff, L. and L. Vaughan, *Co-occurrence Matrices and their Applications in Information Science: Extending ACA to the Web Environment.* Journal of the American Society for Information Science and Technology, 2006. **57**(12): p. 1616-1628.

[22] DiMaggio, P.J. and W.W. Powell, *Tie Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields.* American Sociological Review, 1983. **48**(2): p. 147-160.

[23] Creswell, J.W. and V.L. Plano Clark, *Designing and Conducting Mixed Methods Research*. 2007, Thousand Oaks: Sage.